

Supplement to the response to the reviewer comments, August 9, 2021

Figure A

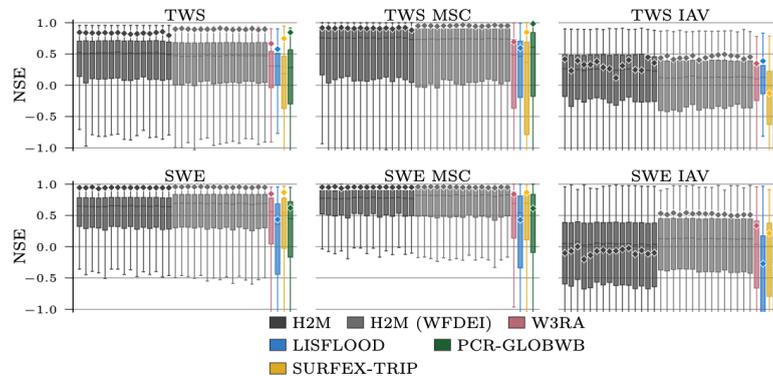


Figure D1. Global and local grid cell level Nash–Sutcliffe model efficiency coefficient (NSE) of the hybrid hydrological model (H2M) and the process-based global hydrological models (GHMs) for the terrestrial water storage (TWS) on top and the snow water equivalent (SWE) on bottom. The gray bars represent the cross-validation runs using the forcings described in Section 2.1.1 (dark, “H2M”), and using the WFDEI forcings as used for in the earth2Observe ensemble (light, “H2M (WFDEI)”). The \diamond -markers show the global (spatially averaged per timestep) model performance, the boxes represent the spatial variability of the cell level performance. The panels show the model performance in respect to the full time-series, the mean seasonal cycle (MSC) and the interannual variability (IAV). Note that for SWE, only grid cells with at least one day of snow are shown, as the NSE is not defined if the observations are constant zero, which would lead to a comparison of different grid cells. The y-axis is cut at -1 due to some large negative NSE values. The metrics are calculated from the complete common time-range from 2003 to 2012. Note that deviations from the numbers reported in Tab. 3 are due to different time ranges.

Figure B

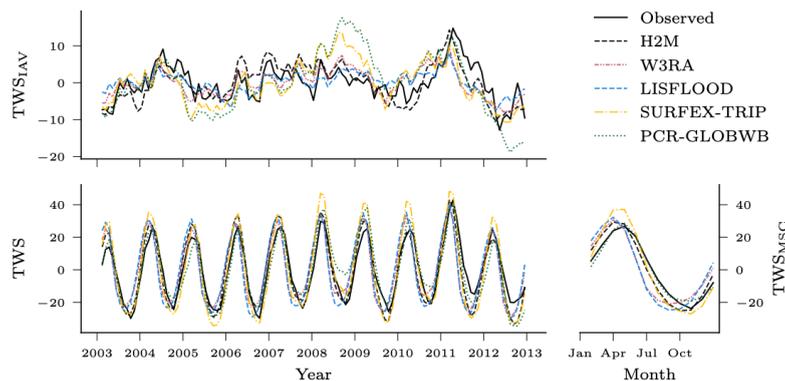


Figure 5. Comparison of the hybrid hydrological model (H2M) and a set of process-based global hydrological models (GHMs) of the terrestrial water storage (TWS), its mean seasonal cycle (TWS_{MSC}) and its interannual variability (TWS_{IAV}) in mm for the global signal. The time-series were aggregated using the cell size weighted mean across all grid cells. The regional time series are shown in Appendix B, Fig. B1.

Figure C

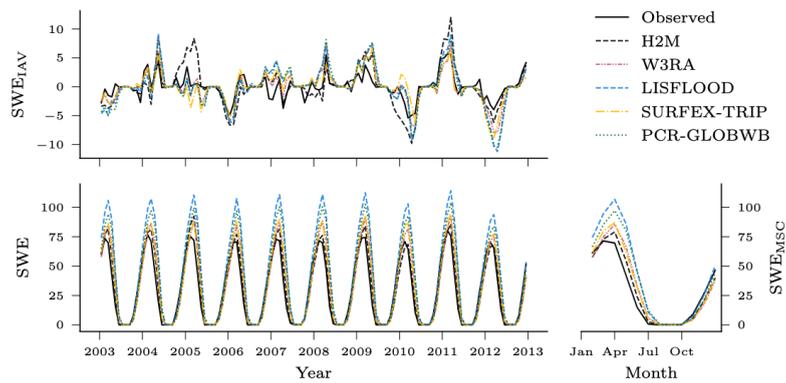
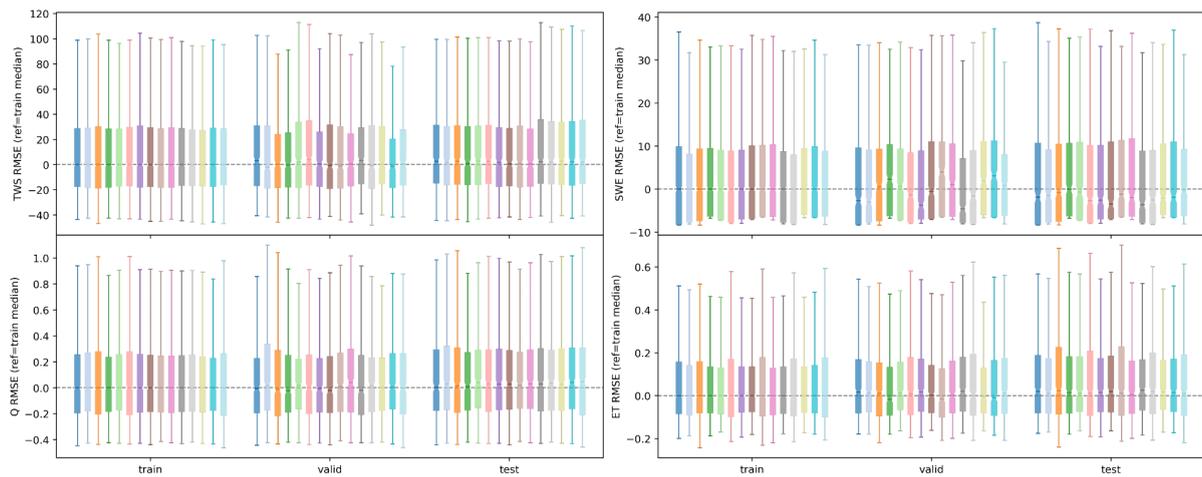


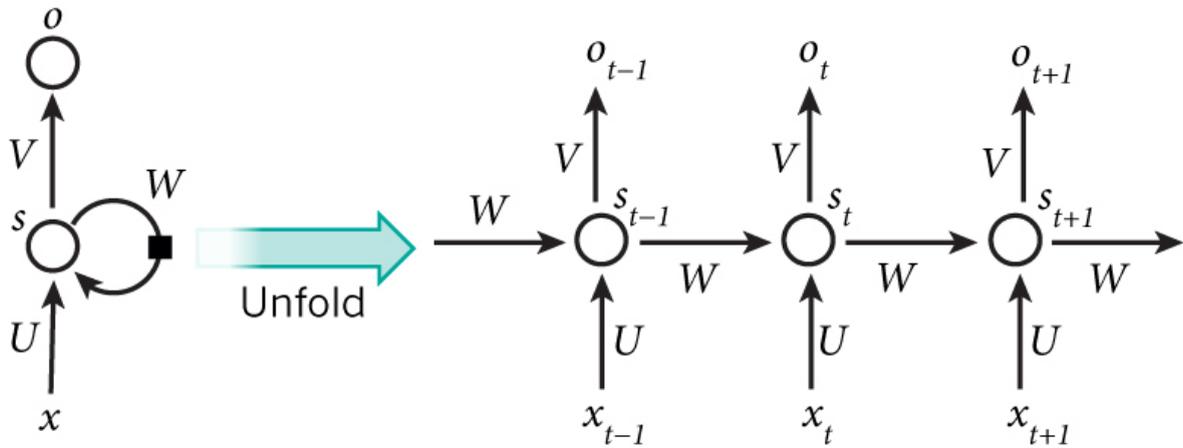
Figure 6. Comparison of the hybrid hydrological model (H2M) and a set of process-based global hydrological models (GHMs) of the snow water equivalent (SWE), its mean seasonal cycle (SWE_{MSC}) and its interannual variability (SWE_{IAY}) in mm for the global signal. The time-series were aggregated using the cell size weighted mean across all grid cells. The regional time series are show in Appendix B, Fig. B2.

Figure D



Generalization error in terms of RMSE. The boxplots represent the spatial variability of the RMSE per cross-validation fold (colors). For each cross-validation fold, the median RMSE of the respective training set was removed, such that the training median is at zero, and the validation and test set boxplots show the RMSE relative to the training set median.

Figure E



Recurrent neural network (RNN) as a loop (left) and unfolded (right). LeCun, Bengio, and G. Hinton, Nature, 2015 / Figure 5.

Table A

Table 4. Global yearly evapotranspiration (ET), runoff (Q), precipitation (Precip.), and storage change (Δ Storage) over the period from 2003 to 2012. The H2M model was forced with the GPCP precipitation product, the other models with WFDEI. The values for H2M and H2M (WFDEI) represent the mean \pm the standard deviation across all cross-validation runs. Values from the common land-mask of all models were considered.

| Model | ET (mm yr ⁻¹) | Q (mm yr ⁻¹) | Precip.* (mm yr ⁻¹) | Δ Storage (mm yr ⁻¹) |
|-------------|------------------------------|-----------------------------|------------------------------------|--|
| H2M | 564 \pm 6.7 | 274 \pm 6.5 | 860 | 21.4 \pm 1.1 |
| H2M (WFDEI) | 553 \pm 6.0 | 285 \pm 6.5 | 851 | 12.9 \pm 1.0 |
| W3RA | 515 | 332 | 851 | 2.5 |
| LISFLOOD | 468 | 397 | 851 | -14.3 |
| SURFEX-TRIP | 552 | 296 | 851 | 2.3 |
| PCR-GLOBWB | 504 | 348 | 851 | -1.3 |

* GPCP for H2M, else WFDEI.